Treebanking in the Language of Thought

Treebanking in the Language of Thought

Mark Steedman
with Tom Kwiatkowski, Emily Thomforde, Sharon Goldwater, Luke Zettlemoyer,
Mike Lewis, *and others*

November 29, 2012

Outline

- I: Language Acquisition from real child-directed data as learning a parser for the Language of Thought (Kwiatkowski *et al.* 2012).

- II: Semiautomatic generation of treebank-size datasets for such learning using the Language of Thought.

- III: Putting the Two Together for Applications in SMT

I: Child Language Acquisition as Learning a Semantic Parser

- Thompson and Mooney (2003); Zettlemoyer and Collins (2005, 2007); Wong and Mooney (2007); Lu *et al.* (2008); Börschinger *et al.* (2011); Liang *et al.* (2011) generalize the problem of inducing parsers from language-specific treebanks like the Penn Treebank to that of inducing "Semantic Parsers" from paired sentences and unaligned language-independent logical forms.
  - The sentences might be in any language.
  - The logical forms might be database queries, λ-terms, robot action primitives, etc.

Semantic Parsers

- This is a harder problem than inducing a standard treebank parser for a given language: in the worst case, we would have to consider all possible pairings of all possible substrings of the sentence with all possible subtrees of the logical form.

- Most of these programs invoke (English) language-specific assumptions.

Inducing Semantic Parsers: GeoQuery

- Kwiatkowski *et al.* (2010, 2011) have applied a more language-general approach to the problem of inducing multilingual grammars from the GeoQueries database of sentence meaning pairs (Thompson and Mooney, 2003):
  - Which states border states through which the mississippi traverses?
  - $\lambda x.\exists y[state(x) \wedge state(y) \wedge loc(mississippi\_river, y) \wedge next\_to(x, y)]$
- GeoQuery is all about *wh*-dependency
- Model is discriminative (log-linear), learned by batch mode inside-outside EM using stochastic gradient descent, iterated, evaluated by 10-fold cross-validation.
- Learning is accelerated by inititialization with Giza++ alignment between strings and logical forms.

Inducing CCG Semantics Parsers: GeoQuery 250

- % of unseen test sentences parsed correctly by induced grammars:

	UBL-s	λ-WASP	Lu08
English	**81.8**	75.6	72.8
Spanish	**81.4**	80.0	79.2
Japanese	**83.0**	81.2	76.0
Turkish	**71.8**	68.8	66.8

- This is done without the language-specific engineering of the other approaches. Constraints on splits are universal (e.g. ATB, A-over-A, semantic-syntactic types mapping).
- See Kwiatkowski *et al.* (2011) for effect of factored lexical generalization, and competitive results on the much harder ATIS travel bookings dataset.

Language Acquisition from Real Child-Directed Utterance

- The child's problem is similar to the problem of inducing a semantic parser (Siskind 1992; Villavicencio 2002, 2011; Buttery 2006).
  - Children too have unordered logical forms in a universal language of thought, not language-specific ordered Penn Treebank trees.
  - So they too have to work out which words (substrings) go with which element(s) of logical form, as well as the directionality of the syntactic categories (which are otherwise universally determined by the semantic types of the latter).

A word may correspond to any substructure of the logical form.

Child and Computer Language Development

- Children do not seem to have to deal with a greater amount of illformedness than that in the Penn Treebank (MacWhinney 2005).
  - But they need to learn huge grammars.
  - They are faced with contexts which support irrelevant logical forms.
  - They need to be able to recover from temporary wrong lexical assignments.
  - And they need to be able to handle serious amounts of lexical ambiguity.
- Geoquery has a tiny grammar, and is so unambiguous that Zettlemoyer and Collins (2005) were able to get away with nothing but a lexical model.

The Statistical Problem of Language Acquisition

- All these programs use a grammar formalism called Combinatory Categorial Grammar (CCG, Steedman 2000, 2012), about which I'll say almost nothing in this talk except that CCG is fully lexicalized and "nearly context free".
- It follows that the task that faces the child is simply to learn the categorial lexicon, given only:
  - (Probably ambiguous, possibly somewhat noisy) sentence-meaning pairs,
  - A universal combinatory mechanism for Syntactic Projection from the lexicon, and
  - A one-to-many mapping from lexical semantic types such as $e \rightarrow t$ to the set of all universally available CCG lexical syntactic types $S\backslash NP, S/NP$, etc..
- Once the lexicon is learned, CCG will handle all constructions, including unbounded dependencies such as relative clauses, for free.

A True Story

- Consider an adult-accompanied child at Piagetian Stage VI who has yet to learn her first word of such a grammar. She encounters a dog, and shows intense interest. Later, she encounters some *more dogs*. The adult observes the child's reaction, and says "MORE DOGGIES!."

- For a few days, the child uses some approximation to the word "doggies" with the meaning *more*, before correcting, apparently happily violating the subset principle.

- We can assume that the child has already learned some phonological regularities of the language by *un*supervised learning, and in particular is in a position to consider the possibility that the utterance consists of more than one word. (Saffran *et al.* 1996; Mattys *et al.* 1999).

More Doggies

- The child thinks: $more'dogs'$

- The Adult says: "More doggies!"

- Given the string "more dogs" paired with the logical form $more'dogs'$, and a mapping from semantic types onto syntactic type like S, NP, $S\backslash NP$ etc., the child can use the universal **BT**-based combinatory rules of CCG to generate

  – all possible syntactic derivations, pairing
  – all possible decompositions of the logical form with
  – all possible word candidates

- Learning a language is just learning its lexicon and a parsing model.

The Derivations

- CCG permits just three derivations for the new utterance "More doggies" , as follows:

(1) a.

$$\dfrac{\dfrac{\textsc{More}}{NP/N : more'_{((e,t),e)}} \quad \dfrac{\textsc{doggies}}{N : dogs'_{(e,t)}}}{NP : more' dogs'_e}{>} \quad !$$

b.

$$\dfrac{\dfrac{\textsc{More}}{N : dogs'_{(e,t)}} \quad \dfrac{\textsc{doggies}}{NP\backslash N : more'_{((e,t),e)}}}{NP : more' dogs'_e}{<} \quad !$$

c.

$$\dfrac{\textsc{More doggies}}{NP : more' dogs'_e} \quad !$$

The Algorithm

- Variational Bayes (cf. Sato 2001; Hoffman *et al.* 2010)

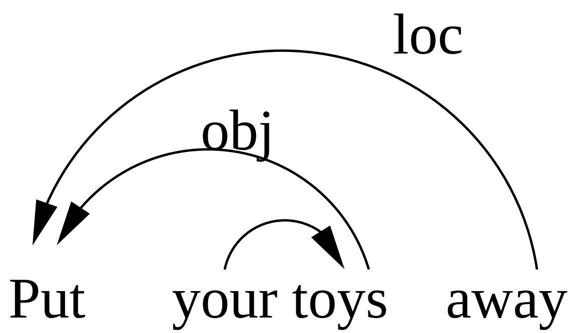
- Incremental Two-stage Expectation/Maximisation Algorithm

- The intuition:
  - Compute the probabilities of all analyses of the new sentence on the basis of the previous model.
  - Update the model on the basis of weighted counts of events in the new sentence.

- We use a Dirichlet prior.

Experiment: Learning from CHILDES Data (Kwiatkowski *et al.* 2012)

- Part of the CHILDES corpus ("Eve") is annotated with dependency graphs.
- These are English-specific.



- We can ignore linear order and treat them as impoverished logical forms.
- In fact we automatically map them into equivalent λ-terms.

Limitations of CHILDES Data

- The resulting pseudo-logical forms are still partly lexically English-specific.

- We will learn constructions that are *less analytic* in other languages than English as multi-word items.

- Milou traverse la rue à la course! *(Milou runs across the road!)*

(2)
$$\dfrac{\dfrac{\text{Milou}}{S/(S\backslash NP) : \lambda p.p\ milou'} \quad \dfrac{\text{traverselarueàla}\textsc{course}}{S\backslash NP : \lambda y.run'(across'road')y}}{S : run'(across'road')\ milou'}{>} \quad !$$

- CHILDES isn't annotated in the Language of Thought accessed by the child.

Using CHILDES Data

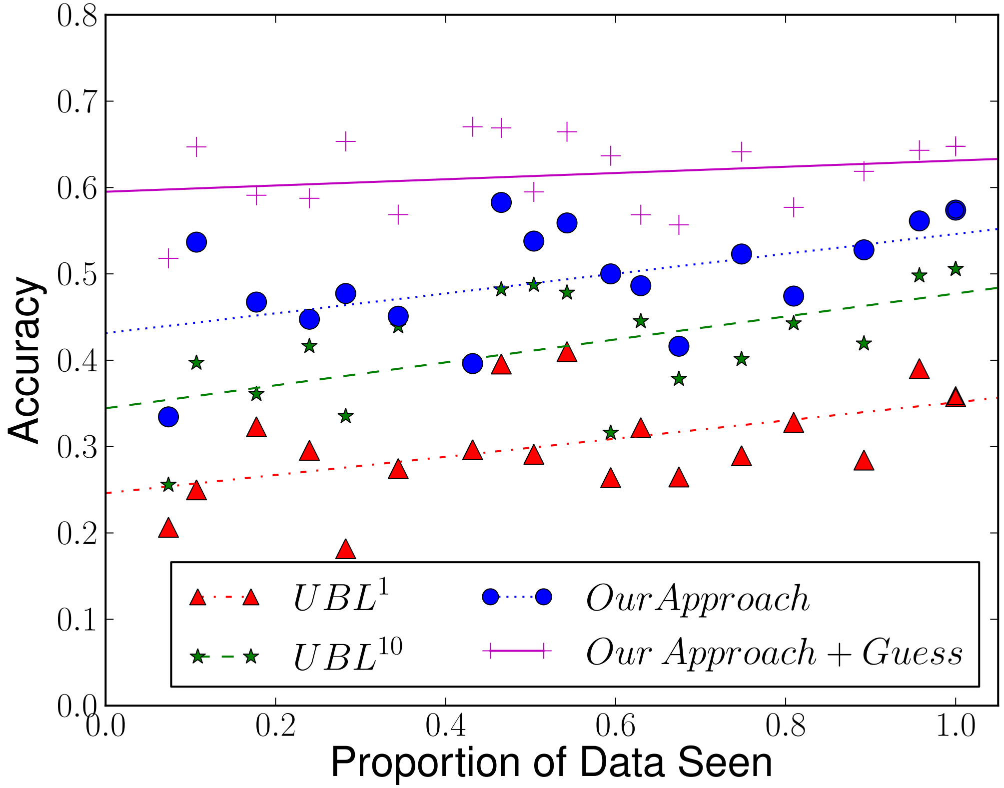
- Nevertheless, we can learn any construction in any language that is *more analytic* than English.

- Range tes jouets! (*Put away your toys!*)

(3)
$$\dfrac{\dfrac{\text{RANGE}}{S/NP : \lambda x.put'away'x\ you'} \quad \dfrac{\text{TES JOUETS}}{NP : toys'}}{S : put'away'toys'you'}{>} \quad !$$

School of informatics

Results (Kwiatkowski *et al.* 2012 EACL)



Results (Kwiatkowski *et al.* 2012 EACL)

- Following Alishahi and Stevenson (2008), we train on chronological sections $1-n$ and test on $n+1$, with up to 6 nearby logical forms as distractors.
- We see steady learning for both this program and Kwiatkowski *et al.* (2010) (the latter with the Giza++ alignment initialization turned off and run for 10 iterations over $1-n$.)
- The present program learns around a steady 8% better than the latter State-of-the-Art semantic parser inducer.
- Even with Giza alignment it is around 1% better.

Results (Kwiatkowski *et al.* 2012 EACL)

- Absolute accuracy of all systems is low because we can only learn from 33% of Eve, excluding stuff like "MMMMMM!" and "DOGGIE DOGGIE DOGGIE!"

- The Eve corpus is also a tiny proportion of what the child has to work with, so test-on-$n+1$ is very brutal.

II: Scaling UBL to MT and Big Data

- We need more datasets! (RoboCup commentaries? Call centers? Robot Navigation?)
- The main problem is (lack of) productivity. (There are only around 600 distinct utterances in RoboCup, unlike GeoQuery.)
- The "dungeons and dragons" scenario of Chen and Mooney (2011) and the related robot navigation domain of Tellex *et al.* (2011) are iteratively productive, but not recursively productive.
- What we need is something like the Penn Treebank annotated with logical forms in the Language of Thought that the child accesses. . .
- . . . also known in Machine Translation as a universal "interlingua".

Aside: Why Did Interlingual MT Fail?

- The language of thought that we are born with has emerged from social interaction and evolutionary processes, and seems quite unlike naive logicist semantics like that in corpora like GeoQuery.
- For example, the logical form that the child brings to learning from "You LIKE the doggies" is likely be something more like $give'pleasure'you'dogs'$, so that the lexical entry for "like" of type $(e,(e,t))$ is the following, exhibiting the same "quirky" relation between (structural) nominative case and an underlying (inherent) dative role that Icelandic exhibits morphologically for the corresponding verb:

(4) like := $(S\backslash NP)/NP : \lambda x \lambda y.give'pleasure'\ y\ x$

Aside: Why Did Interlingual MT Fail?

- With all due respect to cognitive linguists, the obscurity of this hidden semantics, together with the degree of ambiguity in natural grammars, probably makes it impossible to write the grammar for the Language of Thought.

- We can't even get the semantics of prepositions right!

- But if we can get the linguists to tell us all the categories that have ever been observed to be grammaticalized in any language, then maybe machine learning can do the rest.

Case A:Synthetic Motion Verbs

- For example, the French synthetic treatment of motion verbs arguably shows up in Thai and Chinese:

Les enfant	entrent	dans la piece	à la course
the children	enter	into the room	at the run

"The children run into the room"

dek[1]	wing[2]	khaw[2]	ho'ng[2]	(Thai: Diller (2006):167
child	run	enter	room	

"The children run into the room"

háizi-men	pǎo-	-jìn	wūzi	(Mandarin: Li and Thompson (1981))
children	run-	-enter	room	

"The children run into the room"

Case B: Navajo Morphology

- Navajo is comparatively poorly off for nouns. Many nouns for artefacts are morphological derivatives of verbs denoting a protypical affordance.
- For example, "teacher" is *bá'ólta'í*, a morphological compound made up as follows:

b	á	'	ó	l	ta'	í	(Faltz 2000:48n10).
3	for	UNSPEC	3-SUBJ	CL	read	NOM-WH	

  "one for whom one reads"

- Similarly, "towel" is *bee 'ádít'oodí*, glossed as "thing with which one wipes oneself", and "towelrack" is *bee 'ádít'oodí bąąh dah náhidiiltsos*—roughly "what the flat thing with which one wipes oneself is repeatedly put on".

Navajo Classifiers

Classifier+Stem	Label	Explanation	Examples
-'ą	SRO	Solid Roundish Object	bottle, ball, boot, box, etc.
-yí	LPB	Load, Pack, Burden	backpack, bundle, sack, saddle, etc.
-ł-jool	NCM	Non-Compact Matter	bunch of hair or grass, cloud, fog, etc.
-lá	SFO	Slender Flexible Object	rope, mittens, socks, pile of fried onions, etc.
-tą	SSO	Slender Stiff Object	arrow, bracelet, skillet, saw, etc.
-ł-tsooz	FFO	Flat Flexible Object	blanket, coat, sack of groceries, etc.
-tłéé'	MM	Mushy Matter	ice cream, mud, slumped-over drunken person, etc
-nil	PLO1	Plural Objects 1	eggs, balls, animals, coins, etc.
-jaa'	PLO2	Plural Objects 2	marbles, seeds, sugar, bugs, etc.
-ką́	OC	Open Container	glass of milk, spoonful of food, handful of flour, et
-ł-ţí	ANO	Animate Object	microbe, person, corpse, doll, etc.

Do Navajo Classifiers Reflect The Language of Thought?

- Should we include them in our universal meaning representation?
- Such noun classifiers show up in a number of forms across languages.
- Allan (1977) claims that classifers for “animates, food, saliently one-dimensional objects, saliently three-dimensional objects, and for a residual class of general inanimates” recur in all types of classifier language, while a Borgesian list of markers for sharp or hump-shaped objects, trees, gender, mass/count objects, boats, saliently two-dimensional objects, and seed-like objects recur with less generality.
- This seems like the kind of messy mixture of semantics, pragmatics and just plain perversity that we should ignore in the semantics, and just leave the machine learning to sort out.

Case C: Light Verbs and Adverbial Projections

- More interestingly, clauses seem to have a number of semantically transparent "functional projections", like the tensed verb phrase.

- Cinque (1999, 2010) identifies 32 tenses, moods, aspects and voices (TMAV) that are implicitly or explicitly distinguished in a large number of languages, and which attract different classes of morphemes, light verbs, and/or adverbs.

- Cinque's criteria are GB-syntactic, and assume one node per projection and one projection per adverb-type.

- These distinctions seem much more like syntax and semantics as usual (cf. Bittner 2011).

Primitive Verbs

- In particular, they are claimed to form a semantically determined fixed hierarchy, as reflected in head-word order or order of morphological affixation, much like the elements of NPs such as *These five fat boys*.
- For example, there are dialects of English, including one spoken not fifty miles from Edinburgh, that realize "I assume he might be able to go" as a multiple modality "He'll might could go" (Brown 1991).
- In such dialects. the reverse "*He'll could might go" or "*He might will could go" are disallowed.
- Thus we can (cautiously) assume that evidential modality (inference) outscopes epistemic modality (possibility) outscopes root modality (capability and volition).

Linguistic Cartography

- (5) $$\begin{aligned}\text{walks} := S\backslash NP : \lambda y.M_{illocution}\, \mathit{declarative}(M_{mirative}\ \_\ (M_{evidential}\ \_(M_{epistemic}\ \_(\\ T_{remote} -(T_{future} -(M_{realis} +(M_{alethic}\ \_(\\ A_{habitual}\ \_(A_{delayed}\ \_(A_{predispositional}\ \_(A_{repeat}\ \_(A_{frequency}\ \_(\\ M_{root}\ \_(A_{celerative}\ \_(T_{anterior}\ \_(A_{terminative}\ \_(A_{continuative}\ \_(\\ A_{continuous}\ \_(A_{retrospective}\ \_(A_{durative} +(A_{progressive}\ \_(A_{prospective}\ \_(\\ A_{inceptive}\ \_(M_{obligatory}\ \_(A_{frustrative}\ \_(A_{conative}\ \_(\\ A_{sgCompletive}\ \_(A_{plCompletive}\ \_(V_{voice}\ \mathit{active}(A_{celerative}\ \_(A_{inceptive}\ \_(\\ (A_{iterative}\ \_(A_{frequency}\ \_(\textbf{walk}))))))))))))))))))))))))))))))y)))))\end{aligned}$$

  (adapted from Cinque 1999)
- Reflexivity, Case, Agreement, etc. don't seem to belong in logical form, though they are important to the parsing model).
- The place of Causativity, Distributivity, and Topicality is unclear to me.

Adverbs

- (6) John fortunately alledgedly probably now perhaps possibly usually finally wontedly again often willingly quickly already no longer still always soon briefly almost suddenly obligatorily often **walks** repeatedly early completely willfully in vain. $M_{illocution}\ declarative(M_{mirative}\ fortunate(M_{evidential}\ alleged(M_{epistemic}\ probable($
  $T_{tense}\ now(M_{realis}\ +(M_{alethic}\ possible($
  $A_{habitual}\ usual(A_{delayed}\ final(A_{predispositional}\ wonted(A_{repeat}\ again(A_{frequency}\ often($
  $M_{root}\ willing(A_{celerative}\ quick(T_{anterior}\ already(A_{terminative}\ no-longer(A_{continuative}\ still($
  $A_{continuous}\ always(A_{retrospective} just(A_{durative}\ brief(A_{progressive}\ \_(A_{prospective}\ almost($
  $A_{inceptive}\ sudden(M_{obligatory} obligatory(A_{frustrative}\ in-vain(A_{conative}\ willful($
  $A_{sgCompletive}\ complete(A_{plCompletive}\ \_(V_{voice}\ active(A_{celerative}\ early(A_{inceptive}\ \_($
  $(A_{iterative}\ repeated(A_{frequency}\ often(\textbf{walk}))))))))))))))))))))))))))))john))))$

  (adapted from Shlonsky 2010)
- Cinque (2010) proposes a similar hierarchy for nominal modifiers.

A Semantic Reconstruction of Cartography

- Some of these adverbials are mutually exclusive and/or not clearly ordered in scope.
  - T_{past} and T_{future} are exclusive (unless there is more than one Reichenbachian R per tensed clause.)
  - "?He no longer still smokes" and "?He still no longer smokes" seem equivalent and are only non-contradictory to the extent that there we can accomodate an element of indirect speech ("He still *claims* he no longer smokes").

- Maybe we can simplify this map.

A Semantic Reconstruction of Cartography

- We'll assume a Reichenbachian extension of Davidson (1967), in which adverbs are predicated not only over an event time **E**, but over a reference time **R** and the situation of utterance/cognition **S** (Moens and Steedman 1988; Steedman 1997; Bittner 2011).
- The illocutionary, evaluative, evidential, and epistemic modalities are then endotypical (stackable) sentential modifiers $S_{fin}|S$ predicating over **S**, where $|$ marks a verbal or morphological element of either directionality.
- Tense is then an exotypical (non-stackable) VP modifier $(S_{fin}|NP)|(S|NP)$ predicating relations over **E**, **S** and **R**.

A Semantic Reconstruction of Cartography

- Perfect, progressive, prospective, and habitual aspects are exotypical modifiers $(S_\alpha|NP)|(S_\beta|NP)$ mapping **E** onto various consequent, continual, preparatory and dispositional states.
- Root modals map **E** onto deontic/dynamic states
- The remaining (manner etc.)adverbials are predicated of **E** onto deontic/dynamic states

III: How to Treebank the Language of Thought

- Transform an existing treebank CCG parser via the lexicon, to make the same syntactic types build elaborated logical forms by:
  - Expanding analytic terms to synthetic compounds.
  - Replacing open class word meanings by cluster identifiers (Lin; Lewis; Sima'an)
  - Expanding the predicate argument structure by (default valued) predications for attested functional projections like evidentials, miratives, etc.
  - What about Information Structure?
  - What about anaphora/ellipsis?
- Because of lexical ambiguity in elements like modals, this amounts to semiautomatically reannotating the treebank at the level of the terminals—cf. Farwell *et al.* (2009); Bender (2010); Banarescu *et al.* (2012).

Towards Semantic SMT

- Use the English parser with the same lexicon and the same parsing model to parse the English side of a parallel corpus into enhanced logical forms.

- Use the yield of the parser as a logical form bank for the target language.

- Use UBL to induce a semantic parser on the strings and logical forms of the target language.

- (Maybe better start with short sentences ;@)

Rough Example

- The children apparently ran into the room
  - $\lambda \mathbf{e} \lambda \mathbf{r} \lambda \mathbf{s}.enter(the\ room)(the\ children)\mathbf{e} \wedge perfective\ \mathbf{e} \wedge durative\ \mathbf{e}$ $\wedge collective\ \mathbf{e} \wedge manner(running)\mathbf{e} \wedge \mathbf{e} = \mathbf{r} \wedge \mathbf{r} < \mathbf{s} \wedge declarative\ \mathbf{s} \wedge hearsay\ \mathbf{s}$

- Il parâit que les enfants sont entrés dans la chambre à la course.
  - sont $:= (S\backslash NP)/(S_{ppt}\backslash NP_{plural})$: $\lambda p \lambda y \lambda \mathbf{e} \lambda \mathbf{r} \lambda \mathbf{s}.p\ y\mathbf{e} \wedge \mathbf{e} = \mathbf{r} \wedge \mathbf{r} < \mathbf{s}$
  - entrés $:= (S_{ppt}\backslash NP_{plural})/PP$: $\lambda x \lambda y \lambda \mathbf{e}.enter\ xy\mathbf{e}$
  - à la course $:= (S_v\backslash NP_a)\backslash(S_v\backslash NP_a)$: $\lambda p \lambda y \lambda \mathbf{e}.py\mathbf{e} \wedge manner\ running\ \mathbf{e}$
  - Il parâit que$:= S/S$: $\lambda s \lambda \mathbf{s}.s\ \mathbf{s} \wedge hearsay\ \mathbf{s}$

Observations

- We don't have to get it right first time: this process will be robust.
  - It will take null pronouns in its stride.
  - If we hit a target language which marks a classifier or predicate that we have never heard of, the marker will be learned as an empty operator.

- It should be fairly easy to derive a semantics for such operators from the new corpus, by inspection, where appropriate.

- We can improve the English lexicon on that basis for future unseen languages, in a virtuous cycle.

Conclusions

- One should not underestimate the difficulty of this task.
  - There is still a gap between linguistic description and a real semantics.
  - In the absence of the latter some of the syntactic observations are questionable.
    - In some cases it is not clear whether ordering constraints are grammatical or pragmatic
    - Ambiguity and coercion of semantic types is widespread.
- Nevertheless there seems to be a real possibility of making this work.
- The contribution of linguists (especially semanticists) will continue to be essential.

References

Alishahi, Afra and Stevenson, Suzanne, 2008. "A Computational Model of Early Argument Structure Acquisition." *Cognitive Science* 32:789–834.

Allan, Keith, 1977. "Classifiers." *Language* :285–311.

Banarescu, Laura, Bonial, Claire, Cai, Shu, Georgescu, Madalina, Griffitt, Kira, Hermjakob, Ulf, Knight, Kevin, Koehn, Philipp, Palmer, Martha, and Schneider, Nathan, 2012. "Abstract Meaning Representation (AMR) 1.0 Specification."

Bender, Emily, 2010. "On Achieving and Evaluating Language Independence in NLP." *Linguistic Issues in Language Technology* 3:unpaged.

Bittner, Maria, 2011. "Time and Modality without Tenses or Modals." In Renate Musan and Monika Rathert (eds.), *Tense Across Language*, Tübingen: Niemeyer. 147–188.

Börschinger, Benjamin, Jones, Bevan K., and Johnson, Mark, 2011. "Reducing Grounded Learning Tasks To Grammatical Inference." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1416–1425.

Brown, Keith, 1991. "Double Modals in Hawick Scots." In Peter Trudgill and J.K. Chambers (eds.), *Dialects of English: Studies in Grammatical Variation*, Longman. 74–103.

Buttery, Paula, 2006. *Computational Models for First Language Acquisition*. Ph.D. thesis, University of Cambridge.

Chen, David and Mooney, Raymond, 2011. "Learning to Interpret Natural Language Navigation Instructions from Observations." In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-2011)*. 859–865.

Chiang, David, 2005. "A Hierarchical Phrase-Based model for Statistical Machine Translation." In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, MI: ACL, 263–270.

Cinque, G., 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press, USA.

Cinque, Guglielmo, 2010. *The Syntax of Adjectives*. Linguistic Inquiry Monograph 57. Cambridge MA: MIT Press.

Davidson, Donald, 1967. "The Logical Form of Action Sentences." In Nicholas

Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press. 81–95.

Diller, Anthony, 2006. "Thai Serial Verbs: Cohesion and Culture." In Alexandra Aikhenvald and R.M.W. Dixon (eds.), *Serial Verb Constructions: A Cross-linguistic Typology*. Oxford: Oxford University Press, 160–177.

Faltz, Leonard, 2000. "A Semantic Basis for Navajo Syntactic Typology." In Theodore Fernald and Paul Platero (eds.), *The Athabaskan Languages: Perspectives on a Native American Language Family*, New York: Oxford University Press. 28–50.

Farwell, David, Dorr, Bonnie, Green, Rebecca, Habash, Nizar, Helmreich, Stephen, Hovy, Eduard, Levin, Lori, Miller, Keith, Mitamura, Teruko, Rambow, Owen, Reeder, Florence, and Siddharthan, Advaith, 2009. "Interlingual Annotation of

Multilingual Text Corpora and FrameNet." In Hans Boas (ed.), *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, De Gruyter Mouton. 287–318.

Hoffman, Matthew, Blei, David, and Bach, Francis, 2010. "Online Learning for Latent Dirichlet Allocation." *Advances in Neural Information Processing Systems* 23:856–864.

Kwiatkowski, Tom, Steedman, Mark, Zettlemoyer, Luke, and Goldwater, Sharon, 2012. "A Probabilistic Model of Language Acquisition from Utterances and Meanings." In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*. Avignon: ACL, 234–244.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2010. "Inducing Probabilistic CCG Grammars from Logical Form with Higher-

Order Unification." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: ACL, 1223–1233.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2011. "Lexical Generalization in CCG Grammar Induction for Semantic Parsing." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1512–1523.

Li, Charles and Thompson, Sandra, 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

Liang, Percy, Jordan, Michael, and Klein, Dan, 2011. "Learning Dependency-Based Compositional Semantics." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, OR: ACL, 590–599.

Lu, Wei, Ng, Hwee Tou, Lee, Wee Sun, and Zettlemoyer, Luke S., 2008. "A Generative Model for Parsing Natural Language to Meaning Representations." In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: ACL, 783–792.

MacWhinney, Brian, 2005. "Item Based Constructions and the Logical Problem." In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition. CoNNL-9*. ACL, 53–68.

Mattys, Sven, Jusczyk, Peter, Luce, Paul, and Morgan, Jim, 1999. "Phonotactic and Prosodic Effects on Word Segmentation in Infants." *Cognitive Psychology* 38:465–494.

Moens, Marc and Steedman, Mark, 1988. "Temporal Ontology and Temporal Reference." *Computational Linguistics* 14:15–28.

Saffran, Jenny, Aslin, Richard, and Newport, Elissa, 1996. "Statistical Learning by 8-month-old Infants." *Science* 274:1926–1928.

Sato, Maso-Aki, 2001. "Online Model Selection based on the Variational Bayes." *Neural Computation* 13(7):1649–1681.

Shlonsky, Ur, 2010. "The Cartographic Enterprise in Syntax." *Language and Linguistics Compass* 4:417–429.

Siskind, Jeffrey, 1992. *Naive Physics, Event perception, Lexical Semantics, and Language Acquisition*. Ph.D. thesis, MIT.

Steedman, Mark, 1997. "Temporality." In Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam: North Holland/Elsevier. 895–938.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark, 2012. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Tellex, Stephanie, Kollar, Thomas, Dickerson, Steven, Walter, Matthew, Banerjee, Ashis, Teller, Seth, and Roy, Nicholas, 2011. "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation." In *Proceedings of the 25th National Conference on Artificial Intelligence*. AAAI, 1507–1514.

Thompson, Cynthia and Mooney, Raymond, 2003. "Acquiring Word-Meaning Mappings for Natural Language Interfaces." *Journal of Artificial Intelligence Research* 18:1–44.

Villavicencio, Aline, 2002. *The Acquisition of a Unification-Based Generalised Categorial Grammar*. Ph.D. thesis, University of Cambridge.

Villavicencio, Aline, 2011. "Language Acquisition with Feature-Based Grammars." In Robert Boyer and Kirsti Börjars (eds.), *Non-Transformational Syntax: A Guide to Current Models*, Blackwell. 404–442.

Wong, Yuk Wah and Mooney, Raymond, 2007. "Learning Synchronous Grammars for Semantic Parsing with Lambda Calculus." In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. ACL, 960–967.

Xu, Peng, Kang, Jaeho, Ringgaard, Michael, and Och, Franz, 2009. "Using a Dependency Parser to Improve SMT for Subject-Object-Verb Languages." In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Boulder, CO: ACL, 245–253.

Zettlemoyer, Luke and Collins, Michael, 2005. "Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial

Grammars." In *Proceedings of the 21st Conference on Uncertainty in AI (UAI)*. Edinburgh: AAAI, 658–666.

Zettlemoyer, Luke and Collins, Michael, 2007. "Online Learning of Relaxed CCG Grammars for Parsing to Logical Form." In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*. Prague: ACL, 678–687.